# Discovery of Geospatial Resources

Methodologies, Technologies, and Emergent Applications

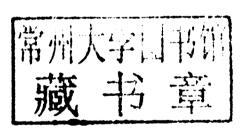
## Discovery of Geospatial Resources:

## Methodologies, Technologies, and Emergent Applications

Laura Díaz Universitat Jaume I de Castellón, Spain

Carlos Granell Universitat Jaume I de Castellón, Spain

Joaquín Huerta *Universitat Jaume I de Castellón, Spain* 





Managing Director: Senior Editorial Director: Book Production Manager: Lindsay Johnston Heather A. Probst Sean Woznicki Joel Gamon

Development Manager: Development Editor:

Mike Killian Erika Gallagher

Acquisitions Editor: Typesetter:

Nicole Sparano

Cover Design:

Nick Newcomer, Lisandro Gonzalez

Published in the United States of America by

Information Science Reference (an imprint of IGI Global)

701 E. Chocolate Avenue Hershey PA 17033 Tel: 717-533-8845 Fax: 717-533-8661

E-mail: cust@igi-global.com Web site: http://www.igi-global.com

Copyright © 2012 by IGI Global. All rights reserved. No part of this publication may be reproduced, stored or distributed in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher. Product or company names used in this set are for identification purposes only. Inclusion of the names of the products or companies does not indicate a claim of ownership by IGI Global of the trademark or registered trademark.

#### Library of Congress Cataloging-in-Publication Data

Discovery of geospatial resources: methodologies, technologies, and emergent applications / Laura Diaz, Carlos Granell and Joaquin Huerta, editors.

p. cm.

Includes bibliographical references and index.

Summary: "This book provides relevant theoretical frameworks and the latest empirical research finding to improve understanding of geospatial discovery methodologies and technologies, as well as techniques to design and deploy geospatial resources in Information Infrastructures"--Provided by publisher.

ISBN 978-1-4666-0945-7 (hardcover) -- ISBN 978-1-4666-0946-4 (ebook) -- ISBN 978-1-4666-0947-1 (print & perpetual access) 1. Geographic information systems--Research. 2. Geospatial data. 3. Spatial analysis (Statistics) I. Diaz, Laura, 1976- II. Granell, Carlos, 1976- III. Huerta, Joaquin, 1965-

G70.212.D64 2012 910.285--dc23

2011048796

British Cataloguing in Publication Data

A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

## Editorial Advisory Board

Anders Friis-Christensen, National Survey and Cadastre, Denmark
Cristiano Fugazza, European Commission – Joint Research Centre, Italy
Michael Gould, Esri, USA
Gobe Hobona, University of Nottingham, UK
Werner Kuhn, University of Muenster, Germany
Miguel Ángel Manso, Universidad Politécnica de Madrid, Spain
Paolo Mazzetti, National Research Council of Italy, Italy
Javier Nogueras, University of Zaragoza, Spain
Victor Pascual, IDEC, Spain
Sven Schade, European Commission – Joint Research Centre, Italy
Laura Spinsanti, European Commission – Joint Research Centre, Italy

## List of Reviewers

Miguel Ángel Esbrí, AtoS, Spain
Joan de la Flor, Science and Technology Park for Tourism and Leisure, Spain
Laura Díaz, Universitat Jaume I, Spain
Carlos Granell, Universitat Jaume I, Spain
Joaquín Huerta, Universitat Jaume I, Spain
Miguel Luaces, University of A Coruña, Spain
Jorge Rocha, Universidade do Minho, Portugal
Mattia Santoro, National Research Council of Italy, Italy
Luis Vilches, Universidad Politécnica de Madrid, Spain

### Foreword

The web is growing in scale and in directions never imagined, making feasible the notion of someday connecting everything to everything. To reach this nirvana, numerous breakthroughs will be required in the automated tagging, indexing, and discovery of all these "things," more adequately termed resources. Human intervention alone, even harnessing all possible "mechanical turks," is not scalable enough to ever reach satisfactory completion of the connection job. Knowledge and, therefore, resources are growing much faster than the natality in any country, and an increasing number of citizens are now generating and publishing, by the push of a button on a smartphone, more resources than they are discovering. The discovery and publication curves are diverging.

This book presents a collection of attempts to push forward the automated resource discovery frontier in the thematic area of geospatial resources: primarily geodata and geo-services. The discovery of geospatial resources, judging fitness for purpose, reformatting on-the-fly if necessary, and then consuming them in order to provide answers to spatial questions ("will it start raining here soon, and if so which is the nearest bus to where I'll be when that happens?") will make possible a whole set of user scenarios that would be totally unfamiliar to the GIS expert of the year 1990. Some would say progress in this area has been slow, as OGC initiatives and other efforts to move toward intelligent geo-resource-based services have been on the drawing board and, indeed, have been demonstrated for at least the past decade, but this is not to say that the geospatial industry has not learned a great deal from these experiments, even from some of the failures, and that today's software is not more functional discovery-wise as well as publication-wise. Some progress has been made, perhaps not spectacular as measured in Silicon Valley terms, but slow, steady progress for sure. Proof is as simple as asking the average person on the street, in the developed world anyway, if they have used virtual globes on the web to view their home or their next holiday destination.

I applaud the editors of this book for bringing together a wide spectrum of perspectives and possible methods. Which methods will win out? I believe the normal Darwinian shakeout will leave standing those which are either simplest (see Ockham's razor) and/or best integrated in current scientific and commercial workflows. It will be interesting to look back at this volume in a decade and judge for ourselves if this has been the case.

Michael Gould Esri, USA & Universitat Jaume I, Spain Michael Gould earned his Ph.D. in Geography (GIS) from the National Center for Geographic Information and Analysis (NCGIA), University at Buffalo. Dr. Gould currently holds the Esri Global Education Manager position. He is also a part-time professor at Universitat Jaume I, Castellón, Spain, where he is one co-creator of an Erasmus Mundus-funded geospatial technologies master's program. Gould's research expertise includes creating Spatial Data Infrastructures (SDI), researching standards-based interoperability with the Open Geospatial Consortium, Inc. (OGC), and developing geospatial Web services.

## Preface

Distributed Geospatial applications, built using standards-based services, are currently deployed on top of Geospatial Information Infrastructures. This has increased the range of distributed geospatial content, services, and applications publicly available in Internet. Although standardization of services and components increases interoperability, the distribution of resources makes their discovery a complex task. There are many aspects to be considered in order to improve the discovery of geospatial resources. For instance, resource descriptions are a crucial aspect: content providers may choose to provide standard metadata elements, annotations, simple user tags, and also choose which formats to use to represent these descriptions. Other aspects such as where to store these descriptions, in terms of using catalogues services, registries or other distributed storage, and when to generate them, may also influence the discovery process. Another aspect to consider is the search engine used to query. In distributed environments, where standards-based components allow resource reusability, the efficiency of discovery often determines the success of application deployment. Therefore, continued investigation into discovery methodologies, technologies, and innovative applications is needed.

This book aims to provide relevant theoretical frameworks and the latest empirical research findings in the field of geospatial resources discovery. The contributors are scientists and professionals, experts in the field, who describe their work and proposed solutions to improve the understanding of geospatial discovery methodology and technology. The following chapters provide a comprehensive overview of common problems and solutions. The reader will find novel proposals for generating resource descriptions, as well as mechanisms to register and search for geospatial resources. The proposed solutions describe techniques to design novel Geospatial Information Infrastructures to improve the interoperability and the visibility of available resources in order to maximize their reusability.

The book is organized in three sections, each composed of works addressing a distinct challenge explicitly exposed in the title of the book. Section 1 refers to Discovery Methodologies, Section 2 treats Discovery Technologies, and Section 3 deals with Emergent Applications.

#### **SECTION 1: METHODOLOGIES**

Improving the discovery of geospatial resources involves designing and assessing discovery strategies in the realm of the application domain. As the quantity and variety of geospatial resources may vary in different domains and scenarios, capturing domain knowledge, and metadata descriptions becomes essential for efficient discovery. Geospatial resources must be provided with metadata descriptions that are up-to-date and coherent with the characteristics of the resource, to enable efficient discovery. Chapters

1-5 explore distinct methodologies for discovery of geospatial resources related to specific application domains, as well as methodologies to generate metadata descriptions to ensure consistency over time and to facilitate the discovery of geospatial resources.

## Chapter 1: Describing and Selecting Collections of Georeferenced Media Items in Peer-to-Peer Information Retrieval Systems

Peer-to-Peer (P2P) systems consist of distributed computers that can act as both clients and servers. Enabling information retrieval of geo-referenced media items in some P2P systems depends highly on the indexing mechanisms and selection methodologies to adequately describe and identify the media collections of the peers. Chapter 1, authored by Daniel Blank and Andreas Henrich, analyses different indexing and selection techniques to perform geospatial k nearest neighbour queries in P2P networks, that is, discovering efficiently the peers with the k closest media items according to a given query location. The chapter demonstrates how geospatial based searches can be performed to find geo-referenced media items distributed over multiple peers.

## Chapter 2: An Assessment of Several Taxonomies of Volunteered Geographic Information

Regarding the generation of resource descriptions, Antony K Cooper, Serena Coetzee, and Derrick G Kourie describe, in Chapter 2, a study which aims to assess several taxonomies related to the concept of user-generated content in general and Volunteered Geographic Information (VGI) in particular. In line with the current trend of leveraging social networks and crowdsourcing platforms, and from the approach of using volunteered contribution of resources by regular citizens, the authors pose the question whether the examined taxonomies can accurately and suitably identify VGI resources. The authors assess five taxonomies from the literature, both subjectively and using formal concept analysis to determine their discrimination adequacy. This chapter includes a discussion on how these taxonomies may eventually facilitate the discovery and integration of VGI resources.

## Chapter 3: New Discovery Methodologies in GIS: Improving the Information Retrieval Process

Chapter 3, written by Nieves Brisaboa, Miguel Luaces, and Diego Seco, provides a thorough review of existing methodologies and techniques in the field of geospatial information retrieval. Based on the current state of the art, the authors propose a system architecture for improving this retrieval. Finally, the work presented in this chapter describes in detail an index construction workflow and the design of processing services to assess queries and user interfaces.

### Chapter 4: Managing Schema Evolution in a Federated Spatial Database System

Chapter 4, authored by Xiaoying Wu et al., presents a framework to effectively manage the schema evolution in federated spatial database systems, which integrate multiple spatial data sources in a distributed fashion. Such spatial data systems are an extension of traditional databases but which include spatial descriptions and relationships between stored entities. In particular, the chapter addresses the issue of ensuring schema consistency by propagating schema changes across multiple databases. This propagation methodology automatically updates the corresponding spatial views and queries over a federated environment, in order to provide a unified data access and query mechanism. The authors demonstrate this technology in a so-called Shared Land Information System.

#### Chapter 5: Automatic Metadata Generation for Geospatial Resource Discovery

Metadata descriptions form one of the pillars of current generation of Spatial Data Infrastructures to provide efficient discovery. The need for the creation of metadata has been identified as an important key to overcome the current lack of metadata available in SDI. Chapter 5, authored by Miguel Angel Manso and Arturo Beltran, highlights the importance of metadata for enabling interoperability and discovery of geospatial resources. It analyzes the process of metadata generation from different points of view. This work provides several examples of metadata generation tools applied to diverse types and formats of geospatial resources.

#### SECTION 2: TECHNOLOGIES

From another perspective, beyond theory and methodological frameworks, technology enablers provide the pragmatic approach to address open issues in the field of geospatial discovery. This section describes trends in technological progress within the geospatial resources discovery domain. Current technological developments in the field of geospatial discovery are focused on semantic technologies, which are central to providing semantic interoperability among geospatial resources. These resources are generally described or annotated by geospatial vocabularies or ontologies. The use of geospatial ontologies and knowledge poses important challenges to achieving geospatial reasoning among related resources for geospatial discovery.

On the other hand, social networks and volunteered geographic information are gaining attention among geospatial users. Contrasting the top-down architecture of SDI, mainstream IT systems, and, in particular, the emerging social network services allow active user participation and are becoming a massive source of dynamic geospatial resources. These emerging technologies provide new paradigms in content provision as well as in semantic annotation of resources. Therefore, in order to enrich current SDI with social information these new technologies must be taken into account. Chapters 6-10 discuss architectures, annotation methods, and applications of semantic discovery of geospatial resources, both in SDI as well as over social network services and VGI content.

### Chapter 6: Discovering Geosensor Data By Means of an Event Abstraction Layer

Environmental monitoring systems are essential to provide geospatial information services that ensure public safety in areas commonly affected by natural disasters. With the increasing number of sensor networks being deployed, these systems address the challenge of handling the continuous streaming of sensor data. In Chapter 6, Alejandro Llaves and Thomas Everding describe how event processing of geosensor data can be helpful for discovery of information over real-time stream flows. Semantic technologies are used to link event patterns, acting as data filters, to domain knowledge representations of reality. This approach provides an event abstraction layer on top of sensor web services with the goal of improving environmental change detection methods.

## Chapter 7: Semantic Enrichment for Geospatial Information in a Tourism Recommender System

Joan De la Flor et al., in Chapter 7, present a novel application called SIGTUR/E-Destination. This application develops a technology to enable a recommender system based on semantically enriched geospatial information. It aims to assist the travel sector with sustainable management of complex tourist destinations. The tourism recommender system relies on artificial intelligence planners. In order to improve the accuracy in discovery criteria and filtering of spatial information, the system leverages tourism domain ontologies, content-based and collaborative techniques with the goal of improving user satisfaction in system recommendations.

## Chapter 8: Semantic Annotation of Geospatial RESTful Services Using External Resources

RESTful services have entered prominently into the information system panorama, competing with more traditional web services thanks to their relative simplicity in modeling communications among participants and their natural suitability for the web. To overcome the current lack of machine-readable descriptions for RESTful services, Victor Saquicela, Luis Vilches-Sánchez, and Óscar Corcho, in Chapter 8, present an approach to automate the semantic annotation of geospatial RESTful services using geospatial taxonomies which are publicly available. The chapter describes a heuristic-based approach to integrate cross-domain ontologies together with other external resources to disambiguate the semantic meaning of the input and output parameters of geospatial RESTful services.

#### Chapter 9: Methodologies for Augmented Discovery of Geospatial Resources

Both information technology and geoscience are dynamic disciplines, which undergo continuous change. Technology enablers must be designed to accommodate existing infrastructures to future changes. Among others, semantic interoperability and the integration of web 2.0 resources present current challenges. Chapter 9, authored by Mattia Santoro et al., describes a solution based on a Discovery Augmentation Methodology to enhancing geospatial information discovery capabilities. The methodology is realized in the design and development of components which, when deployed in Geographic Information Infrastructures (GII), address present and future needs.

## Chapter 10: Data Mining Location-Based Social Networks for Geospatial Discovery

In Chapter 10, Edward Pultar proposes an approach for data mining of location-based social networks. The author describes how to extract descriptions from the social profiles based on data mining techniques to identify distinct profile groups in social networks according to their space, time, and activities, in order to discover information about a place. As a result, this approach returns a list of automatically-generated keywords to describe the place based on local resident input. For demonstration purposes, the author applies web crawling techniques to the CouchSurfing social network to discover profiles centred on four different geographical areas around the world.

#### **SECTION 3: EMERGENT APPLICATIONS**

Technological and scientific progress provides new tools but also generates new use cases to address. The geospatial domain is a horizontal domain that is gaining momentum in many disciplines. Therefore, from the geospatial domain perspective, recent trends lead to more distributed and dynamic systems and multidisciplinary use cases that pose complex challenges. Section 3 illustrates how geospatial technologies and methodologies can be applied in domains beyond more traditional domains such as the environment. Chapters 11-13 provide emerging applications in varied domains such as spatial applications in epidemiological surveillance, urban planning, and sustainability in touristic scenarios, and web-based user interfaces for geospatial resources discovery within SDI.

## Chapter 11: Distributed Geospatial Data Management for Entomological and Epidemiological Studies

Hugo Martins and Jorge Rocha, in Chapter 11, apply geospatial techniques to the context of epidemiological surveillance to help health professionals to better understand spatio-temporal disease patterns. In particular, the authors focus on the study of Bluetongue (BT), an infectious disease of domestic and wild ruminants, to determine its geographical expansion in conjunction with entomological surveillance programs. The authors describe a web-based application based on open-source geospatial components deployed on a thematic SDI node devoted to manage, query, and visualize entomological data.

## Chapter 12: Analysis of Tourist Behavior Based on Tracking Data Collected by GPS

In Chapter 12, Oriol Bernadó et al. analyse mobility patterns in touristic areas to help urban planners in designing plans and actions to improve the competitiveness and sustainability of tourism activities. The authors describe a web application based on open source components and technologies. This application aims to track pedestrians via GPS devices to identify mobility patterns that would serve as inputs for urban planning in touristic areas.

## Chapter 13: User-Friendly Geoportal Interfaces for Geospatial Resource Discovery

In Chapter 13, Victor Pascual presents an assessment of the development of an operational Catalogue Service and its user interface. This study reflects the design criteria followed during a decade to develop a fully operational catalogue. The study reflects the recent evolution of SDI and the trends regarding geospatial resources description and discovery. The author describes how this functionality has to be provided to users by means of useful and friendly interfaces that are capable of hiding complexity.

#### SUMMARY

This book compiles works, which are split into three areas — methodologies, technologies, and emergent applications — to address discovery issues in the field of geospatial information infrastructures. Our goal has been to offer professionals and researchers a general overview of geospatial resource discovery, and to provide documentation to be used for advanced courses as supplemental material on advances in GIS, spatial data infrastructures, and discovery techniques of geospatial resources in distributed and web environments.

Findings may be applied to various disciplines such as information and communication sciences, environmental management, urban planning, sociology, and disaster management. Hopefully, the compiled work contributes to better insight and to support decision-makers concerned with the discovery and management of geospatial information.

Laura Díaz Universitat Jaume I de Castellón, Spain

Carlos Granell Universitat Jaume I de Castellón, Spain

Joaquín Huerta Universitat Jaume I de Castellón, Spain

## **Table of Contents**

Forewordxi
Prefacexii
Section 1 Methodologies
Chapter 1
Describing and Selecting Collections of Georeferenced Media Items
in Peer-to-Peer Information Retrieval Systems
Daniel Blank, University of Bamberg, Germany Andreas Henrich, University of Bamberg, Germany
That cas Territor, enversity of Earner's, Germany
Chapter 2
An Assessment of Several Taxonomies of Volunteered Geographic Information
Antony K Cooper, University of Pretoria, South Africa
Serena Coetzee, University of Pretoria, South Africa
Derrick G Kourie, University of Pretoria, South Africa
Chapter 3
New Discovery Methodologies in GIS: Improving the Information Retrieval Process
Nieves R. Brisaboa, University of A Coruña, Spain
Miguel R. Luaces, University of A Coruña, Spain
Diego Seco, University of A Coruña, Spain
Chapter 4
Managing Schema Evolution in a Federated Spatial Database System
Xiaoying Wu, Curtin University, Australia
Jianhong (Cecilia) Xia, Curtin University, Australia
Geoff West, Curtin University, Australia
Lesley Arnold, Landgate, Australia
Bert Veenendaal, Curtin University, Australia
in the second of

Chapter 5 Automatic Metadata Generation for Geospatial Resource Discovery		
Section 2		
Technologies		
Chapter 6 Discovering Geosensor Data By Means of an Event Abstraction Layer		
Chapter 7 Semantic Enrichment for Geospatial Information in a Tourism Recommender System		
Chapter 8 Semantic Annotation of Geospatial RESTful Services Using External Resources		
Chapter 9 Methodologies for Augmented Discovery of Geospatial Resources		
Chapter 10 Data Mining Location-Based Social Networks for Geospatial Discovery		

## Section 3 Emergent Applications

Chapter 11
Distributed Geospatial Data Management for Entomological and
Epidemiological Studies
Hugo Martins, Universidade Nova de Lisboa, Portugal & Universidade Técnica de Lisboa, Portuga Jorge G. Rocha, Universidade do Minho, Portugal
Chapter 12
Analysis of Tourist Behavior Based on Tracking Data Collected by GPS241
Oriol Bernadó, Science and Technology Park for Tourism and Leisure, Spain
Albert Bigorra, Science and Technology Park for Tourism and Leisure, Spain
Yolanda Pérez, University Rovira i Virgili, Spain
Antonio P. Russo, University Rovira i Virgili, Spain
Salvador Anton Clave, University Rovira i Virgili, Spain
Chapter 13
User-Friendly Geoportal Interfaces for Geospatial Resource Discovery261
Victor Pascual Ayats, Spatial Data Infrastructure of Catalonia (IDEC), Spain
Compilation of References
About the Contributors
Index 210

## **Detailed Table of Contents**

Forewordxi
Prefacexiii
Section 1 Methodologies
Chapter 1  Describing and Selecting Collections of Georeferenced Media Items in Peer-to-Peer Information Retrieval Systems
Chapter 1 presents and analyses indexing and selection techniques for the description and discovery of collections of georeferenced media content in peer-to-peer (P2P) networks to efficiently perform geospatial k nearest neighbour queries.
Chapter 2  An Assessment of Several Taxonomies of Volunteered Geographic Information
Chapter 2 conducts a study aimed at assessing five different taxonomies related to the concept of user generated content in general and volunteered geographic information in particular by using formal concept analysis to determine the adequacy of their discrimination.
Chapter 3  New Discovery Methodologies in GIS: Improving the Information Retrieval Process

Chapter 3 provides a thorough review of existing methodologies and techniques in the field of geospatial information retrieval and describes an index construction workflow and the design of processing services to assess queries and user interfaces.

Chapter 4	
Managing Schema Evolution in a Federated Spatial Database System	56
Xiaoying Wu, Curtin University, Australia	
Jianhong (Cecilia) Xia, Curtin University, Australia	
Geoff West, Curtin University, Australia	
Lesley Arnold, Landgate, Australia	
Bert Veenendaal, Curtin University, Australia	

Chapter 4 presents a framework to effectively manage the schema evolution in federated spatial database systems which integrate multiple spatial data sources in a distributed way, in order to ensure schema consistency by propagating schema changes across multiple data bases.

#### Chapter 5

Chapter 5 highlights the importance of metadata for enabling interoperability and discovery of geospatial resources and analyzes several metadata extraction methodologies and tools applied to diverse types and formats of geospatial resources.

#### Section 2 Technologies

#### Chapter 6

Chapter 6 highlights the importance of sensor web services in environmental monitoring applications and describes how event processing of geosensor data can be helpful for discovery information over real-time streaming of sensor data.

#### Chapter 7

Joan de la Flor, Science and Technology Park for Tourism and Leisure, Spain

Joan Borràs, Science and Technology Park for Tourism and Leisure, Spain

David Isern, Universitat Rovira i Virgili, Spain

Aida Valls, Universitat Rovira i Virgili, Spain

Antonio Moreno, Universitat Rovira i Virgili, Spain

Antonio Russo, Universitat Rovira i Virgili, Spain

Yolanda Pérez, Universitat Rovira i Virgili, Spain

Salvador Anton-Clavé, Universitat Rovira i Virgili, Spain

Chapter 7 presents a novel approach to enable a recommender system based on semantically enriched geospatial information to assist the travel sector in sustainable management of complex tourist destinations.